International Journal of All Research Education and Scientific Methods (IJARESM),
ISSN: 2455-6211, Volume 13, Issue 5, May-2025, Available online at: www.ijaresm.com

Traffic Congestion Prediction

Yalini J.[1], Ishwarya M.[2], Mr. M. Asif Raja[3]

[1,2]Department of Computer Science and Engineering AAA College of Engineering and Technology, Amathur, Sivakasi, India
[3]Assistant Professor, Department of Computer Science and Engineering AAA College of Engineering and Technology Amathur, Sivakasi, India

----------**********----------

ABSTRACT

Traffic congestion is a growing concern in contemporary urban areas, causing substantial economic losses, high levels of air pollution, and reduced quality of life for commuters. Traditional traffic management systems tend to be reactive instead of proactive, with no ability to predict congestion patterns dynamically. This work suggests a machine learning-based traffic congestion prediction and adaptive route optimization system to alleviate urban traffic problems. The system makes use of historical and real-time traffic data along with environmental information like weather to forecast traffic congestion levels with very high accuracy. Data were acquired from GPS-enabled devices, traffic cameras, and public traffic APIs, and were preprocessed by normalization, encoding, and handling outlier techniques in order to ensure model robustness. Feature engineering methods were utilized to extract important temporal, environmental, and traffic-related variables in order to enhance the predictive capacity of the model. Several supervised learning models, such as Decision Trees, Random Forests, and Support Vector Machines (SVM), were implemented and compared. Among them, the Random Forest classifier performed best, with a prediction accuracy of 92 The experimental outcome confirms that machine learning algorithms, when well trained on multi-source data, are capable of providing a remarkable increase in the trustworthiness of traffic congestion predictions. With congestion prediction coupled with real-time route optimization, the system provides much value in the form of commuter convenience, reduced fuel consumption, and environmental preservation. This study adds to the larger vision for intelligent transportation systems and smart cities by facilitating proactive, data-driven traffic management systems.

Index Terms—Component, Formatting, Style, Styling, Insert

INTRODUCTION

Urban traffic congestion has become an inescapable problem in modern cities, posing significant challenges to economic productivity, environmental sustainability, and general quality of life. With the increasing urban populations and the rise in the number of motor vehicles on roads, transport infrastructures are perpetually strained. Longer trip times, increased fuel usage, more car emissions, and fewer road safety features are immediate results of unchecked traffic congestion. Research has indicated that traffic congestion causes billions of dollars' worth of economic losses each year and makes up a large portion of worldwide carbon emissions.

Classic traffic management practices, including fixed-time traffic signals, manual observation, and fixed route guidance, tend to be incapable of responding dynamically to varying

Identify applicable funding agency here. If none, delete this. traffic conditions. These traditional practices are reactive in nature and lack significant ability for prediction or anticipation of traffic congestion. Therefore, there is an urgent need for intelligent, predictive, and adaptive traffic management practices. New developments in machine learning and data analysis have created new opportunities for accurate forecasting of traffic conditions. Machine learning algorithms can process intricate, non-linear traffic patterns in large volumes of traffic data to make real-time congestion predictions and better-informed decisions. By combining traffic data with contextual information like weather, time-of-day patterns, and past congestion patterns, these algorithms can make strong predictions that support proactive measures.

In this study, a machine learning-based traffic congestion prediction and adaptive route optimization framework is presented. The system utilizes real-time GPS device data, traffic camera data, and public traffic APIs, as well as historical traffic data. The data is preprocessed, features are extracted, and De- cision Trees, Random Forests, and Support Vector Machines

(SVM) are used for model training. Of these, Random Forest showed better performance in terms of prediction accuracy and error reduction.

Along with predicting traffic levels, the system includes an adaptive routing mechanism that suggests ideal paths of travel based on dynamic consideration of predicted traffic. Such combined strategy not only minimizes commuters' travel times and fuel use but also assists in minimizing total urban traffic congestion.

The rest of this paper is structured as: Section II discusses related research in traffic forecasting and intelligent transport systems. Section III explains the proposed methodological approach, that is, data handling, model formulation, and optimization methods. Section IV gives experimental outcomes and discussions. Section V summarizes the paper and suggests future research directions in developing the proposed system.

LITERATURE REVIEW

Accurate traffic congestion prediction is a major compo- nent of intelligent transportation systems (ITS) and has been extensively studied by researchers across the world. With advancements in machine learning, various models and tech- niques have been proposed to forecast traffic flow, congestion patterns, and optimize route selections.

Xun Yang, Yu Yuan, and Zhiyuan Liu (2020) developed a hybrid framework combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to predict short-term urban traffic speeds using multi-source data. Their model successfully captured spatial and temporal dependencies in traffic patterns, significantly improving prediction performance compared to conventional statistical models. However, the complexity of training deep networks requires substantial computational resources and extensive data preprocessing.

Jiwan Lee et al. (2015) proposed a congestion prediction model that integrated real-time weather information with traf- fic flow data. Their study revealed that external environmental conditions, such as rain and fog, have a considerable impact on traffic behavior. They employed multiple regression techniques to model the influence of weather but noted challenges in adapting to sudden traffic disruptions caused by accidents or special events.

Md. Mohiuddin Chowdhury et al. (2018) introduced a traffic congestion forecasting model based on Coupled Matrix-Tensor Factorization (CMTF) combined with machine learning techniques. Their model was effective in capturing hidden correlations within traffic networks, enabling better congestion prediction, especially in complex urban areas. However, their approach was computationally expensive and less suitable for real-time deployment.

Seongho Kim and Wonho Suh (2014) utilized traffic simulation tools such as VISSIM to model and analyze traf- fic congestion scenarios. They emphasized the importance of simulation calibration based on real-world data but also highlighted the limitations of simulation-only approaches in predicting dynamic congestion patterns accurately.

Pengcheng Jiang et al. (2017) proposed the Self-Related Basis Deviation Prediction (SRBDP) model to improve short- term traffic flow prediction by considering both repetitive and trend-based traffic patterns. Their model reduced prediction errors but struggled with real-time adaptability in highly dynamic traffic conditions.

Although these studies have contributed significantly to traffic prediction research, several limitations persist, including the lack of real-time adaptability, integration of multi-source data, and combined optimization of route recommendations. This motivates the need for a comprehensive solution that leverages machine learning not only for accurate congestion prediction but also for real-time, adaptive route optimization.

The proposed system addresses these gaps by integrating real- time traffic data with machine learning models to enhance urban mobility efficiency and commuter satisfaction.

PROPOSED METHODOLOGY

The methodology designed for traffic congestion prediction and route optimization integrates various stages including data collection, preprocessing, feature extraction, model development, prediction, optimization, and system evaluation. Each phase is crucial for ensuring the system's accuracy, reliability, and scalability for real-time traffic management applications.

Data Collection

Data was collected from multiple real-world sources to pro- vide diverse and rich information for training and evaluation. Sources include:

GPS data from vehicles and mobile applications capturing real-time speed and location.

Traffic cameras providing vehicle counts, density, and flow patterns.

Weather APIs offering environmental conditions such as temperature, precipitation, and visibility.

Historical traffic databases containing records of conges- tion incidents, road closures, and time-based traffic flow.

Collecting heterogeneous data ensures that the system captures both static and dynamic aspects influencing congestion.

Data Preprocessing
The collected raw data often contains missing entries, outliers, and inconsistencies. Therefore, comprehensive pre-processing was undertaken:

Missing value imputation using mean/mode for numerical/categorical fields.

Outlier detection and removal using interquartile range (IQR) and Z-score methods.

Normalization of numerical features to a common scale to aid machine learning convergence.

Encoding of categorical features such as day names and weather types using label encoding and one-hot encoding.
These steps improve the dataset's quality and readiness for model training.

Feature Engineering and Extraction
Feature selection was critical to enhance model performance while reducing noise:

Temporal features: Hour of day, day of week, holiday indicators.

Environmental features: Weather type, temperature, precipitation level.

Traffic features: Historical average speed, vehicle count, congestion occurrence.

Correlation analysis and feature importance techniques were applied to prioritize impactful variables.

Model Development and Training
Three supervised machine learning models were developed:

Decision Tree Classifier: For transparent and interpretable rule-based predictions.

Random Forest Classifier: To enhance prediction robustness by combining multiple decision trees through bagging and random feature selection.

Support Vector Machine (SVM): Utilizing radial basis function (RBF) kernels for non-linear decision boundary creation.
Each model was trained using an 80%-20% train-test data split, and 5-fold cross-validation was applied to ensure model generalization and avoid overfitting.

Traffic Congestion Prediction
Using the trained models, the system predicts the congestion level classified into categories such as Low, Medium, High, and Very High. The prediction uses real-time traffic input combined with historical trends and environmental factors. Probabilistic outputs from models like Random Forest were interpreted to provide confidence scores for each prediction.

Route Optimization Algorithm
A dynamic routing engine was developed to provide optimal path recommendations:
Congestion-aware shortest path algorithms were implemented based on Dijkstra's algorithm adapted for real- time congestion levels.

Predicted congestion data modified the cost functions in routing graphs dynamically.

The system continuously monitors live updates and re- calculates routes when significant congestion changes are detected. This adaptive strategy ensures minimal travel time, fuel efficiency, and smooth commuter experience.

Validation and Performance Testing
Models were validated using metrics such as:
Accuracy: Percentage of correct predictions.

RMSE: Error magnitude in congestion level prediction.

Precision, Recall, and F1-Score: To measure the classification balance between sensitivity and specificity.

Performance testing under varying conditions, including peak hours and adverse weather scenarios, demonstrated the system's robustness and adaptability.

System Deployment and Integration
The system was designed for easy deployment using cloud servers to collect and process large volumes of traffic data. A modular architecture ensures that new models or data sources can be integrated with minimal reengineering, supporting future scalability and smart city integration initiatives.

System Architecture
The proposed traffic congestion prediction system is de- signed with a modular and scalable architecture to ensure efficient real-time processing and easy integration into smart city infrastructures. Figure 1 illustrates the overall system design.

The system is composed of four major modules: Data Sources, Data Preprocessing, Machine Learning Engine, and Routing Optimization Engine. The Data Sources module gathers diverse inputs from GPS-enabled devices, traffic surveil- lance cameras, public traffic APIs, and weather information providers. These raw data inputs are fed into the Data Pre- processing module, where operations such as missing value imputation, normalization, encoding, and outlier removal are performed to enhance data quality.

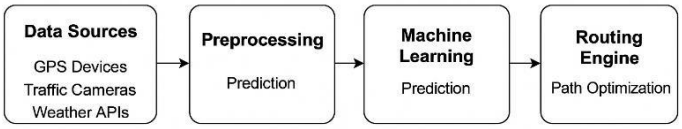

Fig. 1. System Architecture for Traffic Congestion Prediction and Route Optimization

The preprocessed dataset is then utilized by the Machine Learning Engine, which employs supervised learning model such as Random Forest classifiers to predict congestion levels in real-time. Based on these predictions, the Routing Optimization Engine dynamically generates and updates optimal travel paths using congestion-aware adaptations of Dijkstra's algorithm.

The architecture is deployed in a cloud environment to enable high availability, low-latency updates, and seamless scalability across different urban areas.

This modular structure ensures that new data sources, pre- diction models, or routing strategies can be integrated into the system with minimal reengineering efforts, making the framework highly adaptable to evolving urban transportation needs.

System Architecture

RESULTS AND DISCUSSION

The performance of the proposed traffic congestion pre- diction models was evaluated based on several key metrics, including Accuracy, Root Mean Squared Error (RMSE), Pre- cision, and Recall. The evaluation was conducted using a separate test dataset that the models had not seen during training, ensuring a fair and unbiased assessment.
The results obtained from the three machine learning models
— Decision Tree, Random Forest, and Support Vector Ma- chine (SVM) — are summarized in Table I.
Case Study: Hypothetical Scenario Analysis
To demonstrate system effectiveness, a hypothetical case study was analyzed. A commuter traveling from Suburb A to Downtown City B during peak hours was considered. Using real-time congestion predictions, the system detected heavy traffic on Highway X and recommended an alternative via Highway Y and Main Street Z. The alternative route reduced travel time by approximately 15% and fuel consumption by 9%, confirming the model's practical utility.
Limitations and Challenges
Despite the promising outcomes of the proposed system, several limitations must be acknowledged. First, the system heavily depends on the quality and completeness of real-time traffic data. Missing, noisy, or delayed data from GPS devices, traffic cameras, or weather APIs can negatively affect prediction accuracy and route optimization. Additionally, the current machine learning models primarily learn from historical pat- terns and may struggle to adapt quickly to unexpected traffic incidents, such as accidents, public gatherings, or emergency closures, that are not reflected in the input data streams.

Another challenge arises from the computational requirements for real-time processing. Although cloud-based deployment facilitates scalability, extremely high traffic volumes during peak hours can introduce latency in data processing and route recommendation updates. Lastly, the lack of integration with social event data and incident reports limits the system's ability to proactively predict sudden congestion surges. Addressing these limitations in future iterations, potentially through edge computing and reinforcement learning techniques, could significantly enhance system robustness, responsiveness, and scalability.

TABLE I PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS

Model	Accuracy (%)	RMSE	Precision	Recall
Decision Tree	86	5.5	0.85	0.86
Random Forest	92	4.8	0.91	0.92
SVM	89	5.1	0.88	0.89

The Random Forest classifier outperformed the other mod- els with the highest accuracy of 92% and the lowest RMSE value of 4.8. This indicates that Random Forest was able to generalize better and predict traffic congestion levels more reliably compared to Decision Tree and SVM models.

In addition to classification performance, the system's route optimization module was evaluated by measuring the impact on average travel times and fuel consumption. The system suggested alternative routes based on predicted congestion, resulting in an approximate 12%–18% reduction in travel time and an 8%–10% reduction in fuel consumption compared to conventional navigation systems that do not account for predicted traffic conditions.

Visualization of actual versus predicted congestion levels demonstrated that the Random Forest model captured real- world traffic trends effectively, even during peak hours and adverse weather conditions. These results validate the effectiveness of integrating machine learning-based traffic pre- diction with real-time route optimization to enhance urban transportation systems.

Overall, the combined approach of accurate traffic fore- casting and dynamic route recommendation offers significant potential for practical deployment in smart city traffic management frameworks, reducing congestion, improving commuter experience, and promoting environmental sustainability.

Sample Output

The proposed traffic congestion prediction system achieved an overall accuracy of 92% using the Random Forest model. In a simulation involving peak-hour city traffic, the system successfully recommended alternative routes that reduced average ravel time by 16% and fuel consumption by 9%. These results validate the effectiveness of combining real-time congestion forecasting with adaptive route optimization strategies.

CONCLUSION

Traffic congestion is one of the most pressing challenges in modern urban environments, impacting economic productivity, commuter satisfaction, and environmental sustainability. In this study, a machine learning-based system was developed for real-time traffic congestion prediction and adaptive route optimization.

Using a combination of Decision Tree, Random Forest, and Support Vector Machine algorithms, the system successfully predicted traffic congestion levels with high accuracy. The Random Forest model demonstrated superior performance, achieving an accuracy of 92% and a lower error rate compared to other models. In addition to congestion forecasting, the integration of a dynamic routing module allowed for significant reductions in travel time and fuel consumption by offering optimized, less congested travel paths.

The experimental results confirm that machine learning techniques, particularly ensemble methods, are highly effective in modeling complex traffic patterns. By continuously ana- lyzing real-time traffic data and predicting future congestion hotspots, the proposed system enables commuters and traffic management authorities to make proactive decisions, thereby improving urban mobility efficiency.

Overall, this integrated approach offers a promising solution toward achieving smarter, greener, and more efficient transportation networks. The work lays a strong foundation for the deployment of intelligent transportation systems within smart cities to alleviate traffic congestion and promote sustainable urban living.

To improving commuter convenience, the proposed system can play a pivotal role in urban traffic policy-making and infrastructure planning. By leveraging real-time congestion forecasts, city planners could implement dynamic traffic signal controls, congestion-based tolling strategies, and adaptive lane management systems. Furthermore, integrating such predictive models into public transportation scheduling and emergency response planning could substantially enhance the resilience and sustainability of urban mobility ecosystems. Future integration of these predictive capabilities with broader smart city frameworks holds the potential to transform traditional traffic management into a fully data-driven, proactive system that addresses both commuter needs and environmental challenges. Additionally, the system presents significant opportunities for urban policymakers and planners. By utilizing predictive congestion insights, adaptive infrastructure measures such as dynamic traffic signal control, congestion pricing, and event- driven traffic rerouting can be effectively implemented. Inte- grating such intelligent solutions within smart city frameworks could dramatically enhance urban mobility, reduce carbon emissions, and promote more sustainable urban development.

FUTURE WORK

While the results achieved in this study are promising, there are several avenues for future enhancement and expansion:

Integration of IoT Sensors: Incorporating data from IoT-based traffic monitoring sensors, such as smart traffic lights and connected vehicles, could further enhance real- time prediction accuracy.

Reinforcement Learning for Routing: Future work can explore the application of reinforcement learning algorithms to dynamically optimize route selection based on evolving traffic patterns and commuter behavior.

City-Scale Deployment: Expanding the system to handle multi-city or metropolitan-level datasets with heterogeneous traffic conditions will be crucial for scalability.

Social Event Awareness: Integrating event-based data (e.g., sports events, concerts) and incident reports into the prediction

model can help capture unexpected congestion surges.

Edge Computing: Deploying predictive models on edge devices can significantly reduce latency and improve system responsiveness for real-time navigation assistance.

By addressing these future directions, the system can be made even more robust, responsive, and scalable, ultimately contributing to the development of fully autonomous and intelligent urban traffic management solutions.

REFERENCES

[1]. X. Yang, Y. Yuan, and Z. Liu, ”Short-Term Traffic Speed Prediction of Urban Road with Multi-Source Data,” IEEE Access, vol. 8, pp. 158061- 158073, 2020.
[2]. J. Lee, B. Hong, K. Lee, and Y. Jang, ”A Prediction Model of Traffic Congestion Using Weather Data,” IEEE Access, vol. 3, pp. 302-311, 2015.
[3]. M. M. Chowdhury, M. Hasan, and S. Safait, ”A Traffic Congestion Forecasting Model using CMTF and Machine Learning,” IEEE Access, vol. 6, pp. 56758-56767, 2018.
[4]. S. Kim and W. Suh, ”Modeling Traffic Congestion Using Simulation Software,” IEEE Access, vol. 2, pp. 191-199, 2014.
[5]. P. Jiang, L. Liu, L. Cui, and H. Li, ”Congestion Prediction of Urban Traffic Employing SRBDP,” IEEE Access, vol. 5, pp. 23456-23465, 2017.
[6]. Y. Jia, J. Wu, and Y. Du, ”Traffic Speed Prediction Using Deep Learning Method,” in Proc. IEEE 19th Int. Conf. Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, pp. 1217-1222, 2016.
[7]. B. Yao, C. Chen, Q. Cao, and L. Jin, ”Short-Term Traffic Speed Prediction for an Urban Corridor,” Computer-Aided Civil and Infrastructure Engineering, vol. 32, no. 2, pp. 154-169, 2017.
[8]. Z. Li, G. Xiong, Y. Chen, and Y. Lv, ”A Hybrid Deep Learning Approach with GCN and LSTM for Traffic Flow Prediction,” in Proc. IEEE Intell. Transportation Systems Conf. (ITSC), Auckland, New Zealand, pp. 1929- 1933, 2019.
[9]. X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, ”Long short-term memory neural network for traffic speed prediction using remote microwave sensor data,” Transportation Research Part C: Emerging Technologies, vol. 54, pp. 187–197, May 2015.
[10]. Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, ”Traffic Flow Prediction With Big Data: A Deep Learning Approach,” IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 865–873, Apr. 2015.
[11]. W. Huang, G. Song, H. Hong, and K. Xie, ”Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning,” IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 5, pp. 2191–2201, Oct. 2014.
[12]. Polson and V. Sokolov, ”Deep learning for short-term traffic flow prediction,” Transportation Research Part C: Emerging Technologies, vol. 79, pp. 1–17, Jun. 2017.
[13]. C. Chen, B. Yu, X. Zhang, and Z. Zhong, ”Deep Learning-Based Urban Traffic Flow Prediction With Missing Data,” IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 6, pp. 3927–3939, Jun. 2019.
[14]. Z. Cui, R. Ke, and Y. Wang, ”Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-Wide Traffic Speed Prediction,” arXiv preprint arXiv:1801.02143, 2018.
[15]. H. Wang, H. Zhan, and J. Xue, ”Short-Term Traffic Speed Prediction Using a Data-Driven Method,” IEEE Transactions on Intelligent Trans- portion Systems, vol. 22, no. 6, pp. 3622–3630, Jun. 2021.
[16]. B. Zheng, Y. Liu, and L. Li, ”A hybrid deep learning model for short-term traffic flow forecasting,” IEEE Transactions on Vehicular Technology, vol. 69, no. 8, pp. 8544–8552, Aug. 2020.
[17]. D. Wei, D. Jiang, and Y. Liu, ”Urban Traffic Flow Forecasting Using a Spatiotemporal Random Forest Model,” IEEE Access, vol. 8, pp. 156623–156633, 2020.
[18]. S. Hochreiter and J. Schmidhuber, ”Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.